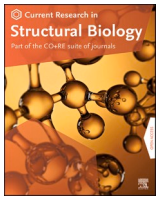
Current Research in Structural Biology 7 (2024) 100142

Contents lists available at ScienceDirect

Current Research in Structural Biology

journal homepage: www.journals.elsevier.com/current-research-in-structural-biology



Back in time to the Gly-rich prototype of the phosphate binding elementary function

Zejun Zheng [a], Alexander Goncearenco [b], Igor N. Berezovsky [a,c,*]

[a] Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, #07-01, Matrix, 138671, Singapore
[b] VantAI, 151 W 42nd Street, New York, NY, 10036, United States
[c] Department of Biological Sciences (DBS), National University of Singapore (NUS), 8 Medical Drive, 117579, Singapore

ARTICLE INFO

Handling Editor: CS Verma

Keywords:
Protein function
Protein evolution
Amino acid chronology
Prebiotic evolution
P-loop NTPases
Rossmann and Rossmann-like
Closed loops
Elementary function
Gly-rich prototype

ABSTRACT

Binding of nucleotides and their derivatives is one of the most ancient elementary functions dating back to the Origin of Life. We review here the works considering one of the key elements in binding of (di)nucleotide-containing ligands – phosphate binding. We start from a brief discussion of major participants, conditions, and events in prebiotic evolution that resulted in the Origin of Life. Tracing back to the basic functions, including metal and phosphate binding, and, potentially, formation of primitive protein-protein interactions, we focus here on the phosphate binding. Critically assessing works on the structural, functional, and evolutionary aspects of phosphate binding, we perform a simple computational experiment reconstructing its most ancient and generic sequence prototype. The profiles of the phosphate binding signatures have been derived in form of position-specific scoring matrices (PSSMs), their peculiarities depending on the type of the ligands have been analyzed, and evolutionary connections between them have been delineated. Then, the apparent prototype that gave rise to all relevant phosphate-binding signatures had also been reconstructed. We show that two major signatures of the phosphate binding that discriminate between the binding of dinucleotide- and nucleotide-containing ligands are GxGxxG and GxxGxG, respectively. It appears that the signature archetypal for dinucleotide-containing ligands is more generic, and it can frequently bind phosphate groups in nucleotide-containing ligands as well. The reconstructed prototype's key signature GxGGxG underlies the role of glycine residues in providing flexibility and interactions necessary for binding the phosphate groups. The prototype also contains other ancient amino acids, valine, and alanine, showing versatility towards evolutionary design and functional diversification.

1. Introduction

The work of physics and chemistry in abiogenesis delivered the so-called primordial soup (Miller, 1953; Miller and Urey, 1959) with basic elements (Goncearenco and Berezovsky, 2015) and consumables (Xie et al., 2015) of the emerging biological world (Romero Romero and Rabin, 2016), establishing building blocks for future biomolecules (Romero Romero and Rabin, 2016; Berezovsky et al., 2000; Eck and Dayhoff, 1966), and introducing basic rules for their functions (Goncearenco and Berezovsky, 2015; Noor et al., 2022). Starting from only few basic chemical reactions acting in the very beginning of biological evolution (Goncearenco and Berezovsky, 2015; Noor et al., 2022; Berezovsky et al., 2017a), the protein function was shaped by the requirements on the thermodynamics and kinetics of reactions, their mechanisms and stereochemistry (Davidi et al., 2018; Riziotis and Thornton, 2022). The protein evolution is a complex hierarchical (Aziz et al., 2016; Dokholyan and Shakhnovich, 2001) process with several major stages (Nath et al., 2014; Trifonov et al., 2001; Siddiq et al., 2017; Trifonov and Berezovsky, 2002, 2003), including those of short prebiotic peptides (Eck and Dayhoff, 1966; Trifonov et al., 2001; Seal et al., 2022) and of ring-like structures (Berezovsky et al., 2000; Berezovsky and Trifonov, 2001a, 2001b) with elementary functions (Goncearenco and Berezovsky, 2010, 2011, 2015; Trifonov et al., 2001; Berezovsky et al., 2003a, 2003b; Alva et al., 2015; Berezovsky, 2019), followed by the stage of small (Goncearenco and Berezovsky, 2015; Romero Romero and Rabin, 2016; Raanan et al., 2020) highly stable functional domains (Zeldovich et al., 2006; Berezovsky, 2003; Trudeau et al., 2016) formed via fusion of respective short genes (Trifonov et al., 2001; Roy et al.,

* Corresponding author. Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, #07-01, Matrix, 138671, Singapore.
E-mail address: igorb@bii.a-star.edu.sg (I.N. Berezovsky).

https://doi.org/10.1016/j.crstbi.2024.100142
Received 30 December 2023; Received in revised form 31 March 2024; Accepted 3 April 2024
Available online 9 April 2024

1999; Sato et al., 1999), and, finally, complemented by the emergence of multi-domain and oligomeric proteins (Goncearenco and Berezovsky, 2012, 2015; Aziz et al., 2016; Trifonov et al., 2001; Aziz and Caetano-Anolles, 2021), protein assemblies, and molecular machines (Berezovsky et al., 2017a, 2017b). The protein function was under the constant selection pressure throughout this evolutionary path, adapting to different extreme environments (Berezovsky, 2011; Berezovsky and Shakhnovich, 2005; Goncearenco and Berezovsky, 2014; Goncearenco et al., 2014; Ma et al., 2010) and their combinations (Goncearenco and Berezovsky, 2014; Goncearenco et al., 2014; Amangeldina et al., 2024) and invoking a number of additional regulatory mechanisms, such as post-translational modifications (Berezovsky et al., 2017a; Johnson and O'Reilly, 1996; Mitternacht and Berezovsky, 2011), allostery (Guarnera and Berezovsky, 2016, 2019a; Tee et al., 2021, 2022), order-disorder transition as triggers of signaling and regulation (Mughal and Caetano-Anolles, 2023; Tee et al., 2020), intermolecular interactions (Aziz et al., 2016) and even fold switching (Dishman et al., 2021). While a contemporary molecular function and protein function, in particular, is a complex and hierarchical phenomenon (Aziz et al., 2016; Aziz and Caetano-Anolles, 2021), the descendants of its basic elements arrived from prebiotic evolution (Siddiq et al., 2017) are still present and play an important role in a diversity of biochemical transformations and other functions (Noor et al., 2022).

2. Function from fragments: from the start in prebiotic world to the emergence and evolutionary design of protein folds and functions

A vast literature describes potential ways and proposes models of the protein domain formation (Goncearenco and Berezovsky, 2015; Romero Romero and Rabin, 2016; Berezovsky, 2003; Berezovsky et al., 2017b) from smaller sub-domains units that existed in primordial folds of the prebiotic world (Heizinger and Merkl, 2021). The origins of symmetry in folds (Broom et al., 2012; Smock et al., 2016), providing a great evolutionary advantage for their functions and regulation thereof (Tee et al., 2022), attracted a specific attention, and received substantial experimental support (Lee and Blaber, 2011). From the fold perspective, on the other hand, the minimal structural units were described in a classical Levitt and Chotiha work back in 1976 (Levitt and Chothia, 1976). Considering less than three dozen of available at that time proteins with solved structures, authors of this seminal work came up with the discovery of three commonly occurring folding units: α-α, β-β, and β-α-β structural patterns. These and only few other basic structural units, "bricks", into which all soluble proteins and major folds can be easily decomposed, are returns of the polypeptide chain or closed loops of nearly standard size of about 25–30 amino acid residues (Berezovsky and Trifonov, 2002a). The polymer nature and evolutionary importance of closed loops, explaining the existence of α-α, β-β, and β-α-β structural patterns (Levitt and Chothia, 1976), are discussed below (Berezovsky et al., 2000, 2017b; Berezovsky and Trifonov, 2001a). Organization of specific folds as a combination of distinct structural units representing closed loops (Berezovsky and Trifonov, 2002a) can be exemplified by the works on the structure of α/β (Sterner and Höcker, 2005; Newton et al., 2017) and β barrels, β propellers (Chen et al., 2011), repeat proteins (Bella et al., 2008) to name a few. These works also reveal hidden evolutionary connections between folds and functions (Goncearenco and Berezovsky, 2010, 2011, 2015; Berezovsky et al., 2017b; Bharat et al., 2008; Farias-Rico et al., 2014; Schneider et al., 2006) determined by their elementary units – descendants of the first functional prototypes (Goncearenco and Berezovsky, 2010, 2011, 2015; Romero Romero and Rabin, 2016; Alva et al., 2015). A number of computational efforts ranging from purely statistical high-throughput analysis of short sequence segments (Kolodny et al., 2021; Nepomnyachiy et al., 2017; Qiu et al., 2022) to classification of proteins with well-described shared motifs (Schaeffer et al., 2016) and discussions of potential scenarios of the fold formation (Goncearenco and Berezovsky, 2012, 2015; Longo et al., 2022a) indicate a non-weakening interest to evolutionary aspects of protein function. Understanding of the discrete structure of modern functional domains provides a strong motivation and foundation for engineering and design efforts (Hocker, 2014; Khersonsky and Fleishman, 2016; Blaber and Lee, 2012; Lechner et al., 2018), which rely on the building of desirable structures and functions from elementary units (Berezovsky, 2019; Yin et al., 2021). Both, fragments of contemporary proteins (Brunette et al., 2015; Huang et al., 2016; Pluckthun, 2015), their reconstructed, simplified, but generic evolutionary prototypes (Berezovsky, 2019; Yin et al., 2021), or *de novo* designed structural units (King et al., 2015; Marcos et al., 2018) can be used depending on the task.

2.1. Basic units of protein structure and function determined by the polymer nature of proteins and shaped by the evolution

Considering diversity of folds and functions, their different evolutionary histories, and several unique external factors that shaped their current appearance and functionality, one may think of generalized consideration of the "basic building block" approach. To this end, a question about generic structure-function unit of proteins, which can be obtained from the analysis of natural proteins and subsequently used in the engineering and design efforts should be raised. Starting from the stability perspective on the minimal unit of protein domain/fold (Berezovsky et al., 1997, 1999), we asked what was the key determinant of the universal basic structural unit of any globular protein regardless of its size, secondary structure composition, or function (Berezovsky et al., 2000). It appeared that the polymer nature of polypeptide chains established a common basic elements of globular proteins (Berezovsky et al., 2000; Berezovsky and Trifonov, 2001a; Berezovsky, 2003) – closed loops or returns of the polypeptide chain of nearly standard size of about 25–30 amino acid residues (Berezovsky et al., 2000, 2017b; Berezovsky and Trifonov, 2001b, 2001c). We showed that this typical size of the polypeptide backbone returns in contemporary proteins fairly agrees with the estimate on the loop size on the basis of the ring-closure probability theory (Shimada and Yamakawa, 1984; Yamakawa and Stokmayer, 1972) and experimentally determined persistence length of polypeptide chains with mixed amino acid composition (Schimmel and Flory, 1967). Another important observation is that the size of closed loops does not depend on the organism (prokaryotic or eukaryotic) hosting the protein or on the specifics of the protein sequence, fold, or secondary structure composition (Berezovsky et al., 2000, 2002, 2017b; Berezovsky and Trifonov, 2001a). All the above corroborates the role of closed loops, or polypeptide chain returns, as universal basic units of globular proteins regardless of the differences in their secondary structure compositions, including the first common folding units described in (Levitt and Chothia, 1976). Indeed, the α-α, β-β, and β-α-β structural patterns represent genuine polypeptide chain returns, closure of which is determined by their own structures, *e.g.* in case of the β-α-β closed loop stabilized by the van der Waals lock between β-strands. The loop closure can also be facilitated by additional interactions, such as van der Waals locks (Berezovsky and Trifonov, 2001b; Koczyk and Berezovsky, 2008) or other interactions with in the context of the overall structure of the protein. Independent and diverse studies extending from the estimates of the ancestral exon sizes (Roy et al., 1999) and centripetal structural modules of proteins (Sato et al., 1999) to observation of the break in power-low pattern of the protein fractal characteristic (Moret and Zebende, 2007), as well as Delaunay (Taylor and Vaisman, 2006) and Voronoi (Angelov et al., 2002) tessellations and other computational experiments (Chintapalli et al., 2014; Yew et al., 2007) provide a clear evidence of the structural and evolutionary relevance of closed loops. Theoretical estimate of the contribution (Berezovsky and Trifonov, 2002b) and the hypothesis on the key role of closed loops in co-translational protein folding (Berezovsky et al., 2001) are strongly supported by experimental works (Ben Ishay et al., 2012; Ittah and Haas, 1995; Orevi et al., 2013) and further substantiated by the modelling of

protein folding dynamics (Bergasa-Caceres and Rabitz, 2018) based on the loop closure as a critical element of the process. Recently, an involvement of closed loops in allosteric signaling in major folds was shown to be a foundation for conservative structure-based allosteric regulation of protein activity, which can be tuned and diversified in the evolution of functional (super)families via mutations and other sequence modifications (Tee et al., 2022).

Another evidence of the involvement of loop closure in cotranslational folding was recently obtained in the statistical analysis of coding sequences (Jacobs and Shakhnovich, 2017), which revealed genes' loci enriched with slowly translated codons associated with cotranslational folding intermediates smaller than a single domain. It was proposed that beneficial pause in synthesis should be separated by a distance similar to the size of ribosome exit tunnel, which can conceal 30–40 residues of nascent protein chain and help some intermediate tertiary structure formation. Analyzing highly conserved regions of rare-codon enrichment, authors, indeed, found a putative translational pause located ~30 residues downstream of a predicted intermediate (Jacobs and Shakhnovich, 2017). Remarkably, both capacity of the ribosome exit tunnel (Samatova et al., 2024; Thommen et al., 2017) and punctuation of coding sequences with rare codons facilitate a formation of folding intermediates with the typical closed loop size. Virtual stability of folding intermediates (Berezovsky and Trifonov, 2001b; Berezovsky et al., 2001) and reduction of the available conformations upon their formation important for obtaining realistic folding times (Berezovsky and Trifonov, 2002b) agree with the role of kinetics in driving of self-assembly according to instructions written in the genetic code (Jacobs and Shakhnovich, 2017).

2.2. Enzymatic function is built form elementary ones – descendants of ancient prebiotic peptides

There is an increasing number of discussions of how everything started some 3.5 billion years ago. Undoubtedly, the amount of sequence/structure data and state-of-the-art experimental techniques make it possible to dig deeper in the evolutionary history and to get fine details of basic "actors" and mechanisms that started the Life. It is also becoming possible to better observe an evolutionary transition to contemporary diversity of structures and functions. It should be noted, however, that remarkable envision on the role of first functional peptides that gave rise to modern folds were made back in 1966 by Eck and Dayhoff on the basis of a handful number of sequences (Eck and Dayhoff, 1966). These functional structures, dubbed "Dayhoff fragments" (Romero Romero and Rabin, 2016), most probably existed and acted on the second stage of protein evolution (Trifonov et al., 2001), preceding formation of the first protein domains/folds upon transition from abiotic (Trifonov and Berezovsky, 2002) to biological evolution (see Fig. 1 in (Berezovsky, 2019) and discussion there). The Dayhoff fragments, in turn, were apparently built from short linear peptides estimated by different authors to be of sizes 3–8 (van der Gulik et al., 2009) or 5–6 (Trifonov et al., 2001) residues, respectively. The amino acid composition of these peptides was biased in favor of the most ancient amino acids (Trifonov and Berezovsky, 2002; Trifonov, 1999), including G, A, V, D, S, and E in order of their appearance documented in the amino acid chronology (Trifonov et al., 2001; Trifonov, 2000). It was shown how compositions of peptides enriched with few amino acid types were related to their first abiotic, then biological functions. For example, search for traces of prebiotic peptides in sequences of protein structure database (PDB) revealed three Aspartic-rich (D-rich) signatures binding

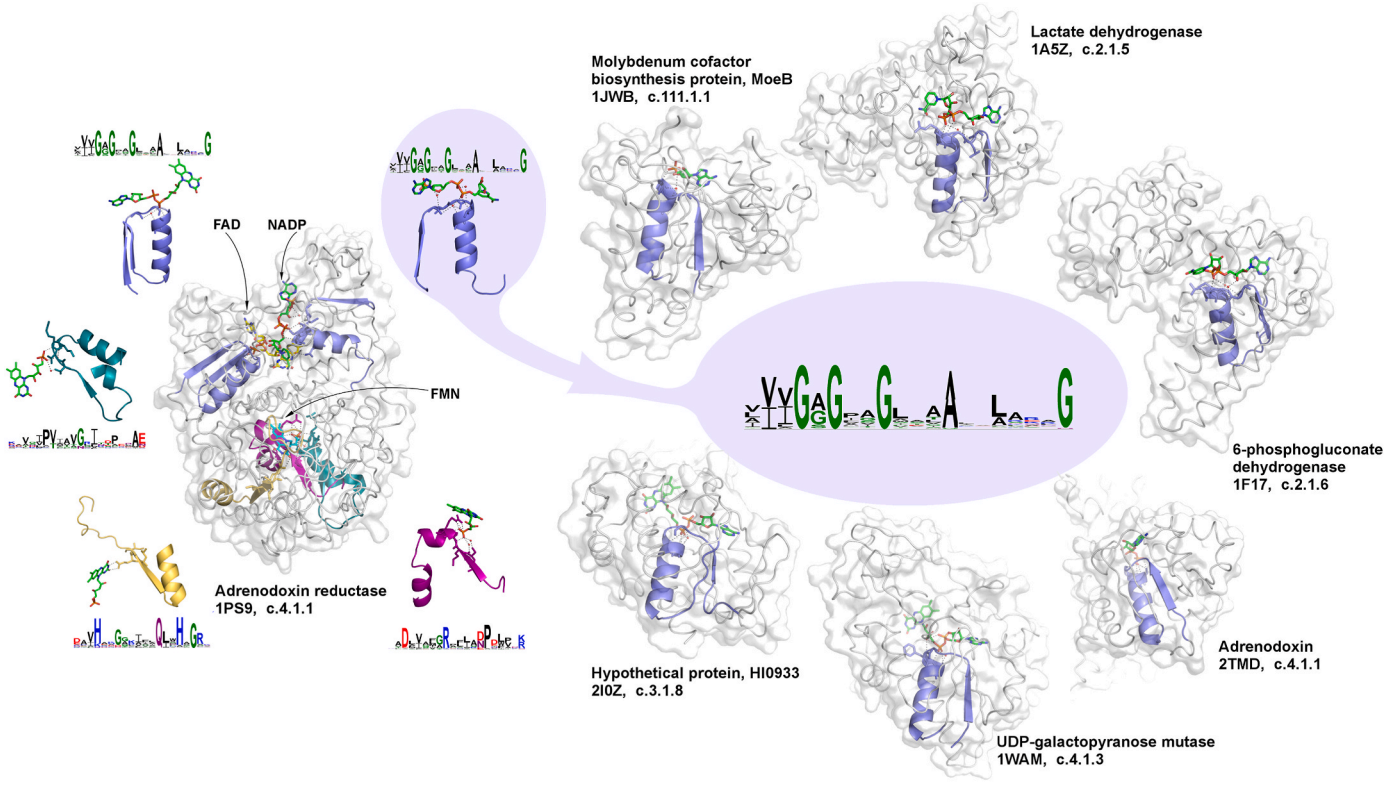


Fig. 1. Sequence/structure signature of the phosphate binding in dinucleotide-containing ligands and its representatives in different biochemical functions. Left panel: Adrenodoxin reductase (PDB ID: 1ps9, nucleotide-binding domain fold) with its elementary functional loops. The glycine-rich motif with a characteristic signature GxGxxG (magenta bulb in the center) binds phosphates in dinucleotide-containing flavin adenine dinucleotide (FAD) and nicotine adenine dinucleotide phosphate (NADP). Set of structures with representatives of GxGxxG signature, showing that it is used in different functional superfamilies and folds: c.111.1.1 is activating enzymes of the ubiquitin-like proteins fold; c.2.1.5 – NAD(P)-binding Rossmann-fold; c.4.1.3 – nucleotide-binding domain; c.3.1.8 – FAD/NAD (P)-binding domain.

mainly Mg^{2+}: D[FY]DGD, DGD[GA]D, and DAKVGDGD, containing a generic motif DGD that was seemingly their common ancestor (van der Gulik et al., 2009). All three signatures are involved in functional manipulation of phosphate groups. This observation prompted authors to conclude that along with a binding of metal ions, interactions with a phosphate could be a very important function in the Origin of Life (van der Gulik et al., 2009). Structure-guided sequence analysis of metal-binding proteins also pointed to the ancient origin of the metal binding D-rich signatures that evolved from binding Mn^{2+} and Fe^{2+} to interactions with other metal ions and hemes (Bromberg et al., 2022). The so-called EF-hand with a signature DxDxDG that binds Ca^{2+} and Mg^{2+} and characteristic helix-turn-helix structural motif is also believed to have a long evolutionary history (Gifford et al., 2007), being presented in folds of all structural classes (Rigden and Galperin, 2004). Analysis of the oxidoreductase superfamily (Raanan et al., 2020) not only confirms early start of the CxxC as a common pattern of signatures that bind different metals (Goncearenco and Berezovsky, 2011). It also establishes a structural link between the metal- and phosphate-binding signatures carried in the same β-α-β structural unit – genuine closed loop and, potentially, one of the first "Dayhoff fragments" capable to safely carry primitive elementary functions. Overall, the repertoire of ancient peptides' functions was seemingly limited to metal and phosphate binding, and the latter is a main topic of this review discussed below. Remarkably, a characteristic Gly-rich signature of the phosphate binding may also serve as a host of another important function. It appears that Gly-rich tracts show an ability for the phase separation and self-assembly (Kar et al., 2021), which could result in evolutionary development of sequence/structures patterns with a structural function of protein-protein interactions. In general, most of the ancient functional signatures are carried by either β-α-β closed loop locked by the van der Waals lock (Berezovsky and Trifonov, 2001b) formed by two β-strands flanking the α-helix, or by the β-turn-α and α-turn-α returns of the protein backbone stabilized within a context of the overall fold. The phosphate (or other ligand) and metal binding signature is typically located in the turn segment of returns and in the turn between first β and α elements of the β-α-β closed loop. In the latter, the second β-strand may contain an additional aspartic acid residue involved in coordination of the metal (Cronet et al., 1995; Laurino et al., 2016), making two most ancient elementary functions to work together.

Moving on to general consideration of the structural role of closed loops in protein evolution (Berezovsky et al., 2000, 2017b; Berezovsky and Trifonov, 2001a) and their involvement in protein function, it was shown that the proteomic code (Berezovsky et al., 2003a) can be derived and used for "spelling" (Berezovsky et al., 2003b) protein sequences of contemporary proteins and annotating their functions. We introduced the operational definition of the elementary functional loop (EFL (Goncearenco and Berezovsky, 2015),) as the unit of structure and function formed by the closed loop (or return of the protein backbone) characterized by the specific sequence with functional residue(s). The EFL form the biochemical function of a protein in combination with other functional residues provided by EFLs of this protein. The evolutionary prototypes of EFLs possess common sequence signatures unraveling deep evolutionary connections between different enzymatic functions (Goncearenco and Berezovsky, 2010), which could not be detected in the analysis of sequences of modern proteins. Therefore, we developed a rigorous statistical approach for derivation of the EFLs prototypes (Goncearenco and Berezovsky, 2011): simplified ancient signatures represented by corresponding EFLs not existing in modern proteins. The fundamental distinction of prototypes lies in their derivation from sequences of unrelated proteins from distant (super)families and even different protein folds (Goncearenco and Berezovsky, 2011). Therefore, contrary to ancestral reconstructions, typically obtained from phylogenetic trees of considered functional superfamilies and giving the origins of functional signatures in this superfamily, the prototypes allow to find their representatives in proteins with no apparent phylogenetic, structural, or functional connections (Goncearenco and Berezovsky, 2015).

2.3. Phosphate binding is one of the cornerstone functions arrived from the prebiotic world

The finding of distant evolutionary connections can be exemplified by the analysis of the evolution of protein function in Archaea (Goncearenco and Berezovsky, 2012), revealing a handful number of functions, including ABC transporters, transferases, helicases, aminoacyl-tRNA synthetases, transcriptional regulators, in which signatures of elementary function were presented (Goncearenco and Berezovsky, 2012). The signatures of the phosphate binding in mono- and dinucleotide containing ligands with generalized patterns GxxGxGK [ST] and GxGxxG, respectively (Goncearenco and Berezovsky, 2010), were among the most omnipresent EFLs. The nucleotide-peptide binding is believed to be the one of the most ancient functions (Goncearenco and Berezovsky, 2010, 2012, 2015) dating back to the Origin of Life (Goncearenco and Berezovsky, 2015; Romero Romero and Rabin, 2016; Trifonov et al., 2001). This intricate connection was established by the emergence and evolution of the triplet code that also determined a temporal order of amino acids (Trifonov, 2000), which, in turn, started interactions between these molecules and exposed them to evolution (Trifonov et al., 2001). In general, studies of related elementary functions in contemporary proteins show that binding of nucleotide-containing ligands is present in a wide spectrum of protein functions (Goncearenco and Berezovsky, 2015; Berezovsky et al., 2017b), as they are indispensable role in corresponding biochemical reaction (Goncearenco and Berezovsky, 2015; Berezovsky et al., 2017b).

Fig. 1 shows an example of a wide representation of the Gly-rich signature (all the signatures here are obtained with the protocol described in Supplementary File and illustrated by Supplementary Figures; see also our earlier works (Goncearenco and Berezovsky, 2015; Goncearenco and Berezovsky, 2010; Goncearenco and Berezovsky, 2011) for further details) of the phosphate binding in dinucleotide-containing ligands in several biochemical functions, as well as three distinct roles of the phosphate binding in one at the same enzyme (left panel). The latter is adrenodoxin reductase, which is the iron-sulphur flavoenzyme required for the metabolism of unsaturated fatty acids (Hubbard et al., 2003). It contains three (!) EFLs of the nucleotide-containing ligand binding: two elementary functions work for binding of FAD and NADP – both are dinucleotide-containing ligands, and one more EFL binds the FMN – the nucleotide-containing ligand. The function of the protein is initiated by the NADPH binding and transfer of the hydride from NADPH to FAD. The latter, then, transfers electrons to FMN, which, being fully reduced, provides a hydride ion to the substrate. This protein is an excellent example of the importance and multiple utilization of the same elementary function in a complex biochemical transformation. At the same time, it also shows how the overall function of a protein can be formed from several different elementary ones (see, for example, magenta, yellow, and green EFLs and their sequence signatures in the left panel). The right panel on the other hand, shows representation of the dinucleotide-containing binding EFL with generic signature GxGxxG in other proteins with different functions. Noteworthy, it is quite common to find several EFs comprising over biochemical function of the protein exemplified in Fig. 1. Another interesting case that we explored based on the profile/prototype reconstruction includes detection of structural EF working in the formation of the homodimer interface (Lasry et al., 2012). Starting from the original task to derive and describe EF of Zn binding (Kambe et al., 2021) formulated by experimental group, we found and derive unknown profile of another EF that provides stabilization of the homodimer. As a result, original hypothesis and conclusions of previous experiments were reconsidered, leading to the inference and demonstration of the critical role of the homodimer formation for efficient transmembrane Zn^{2+} transfer (Lasry et al., 2012).

Because of the omnipresence of the phosphate binding elementary

functions, corresponding sequences and their specific function-related characteristics became a subject of active studies and discussions (Moller and Amons, 1985; Saraste et al., 1990) soon after it was first discovered (Walker et al., 1982). While it started from the originally proposed simplified signature GxxxxGK[ST] (called the P-loop or Walker A motif) obtained on a very limited data (Moller and Amons, 1985; Walker et al., 1982), already a decade later the function and consensus motifs of nine signatures for both nucleotide- and dinucleotide binding were considered (Traut, 1994). Presence of structurally conserved P-loop in widely varying functions, such as muscle contraction in myosin, signal transduction in G protein, phosphoryl transfer in adenylate and guanylate kinases, glutathione synthases, nucleotidyl transferases, oxidoreductases, elongation factors were appreciated and discussed already in mid-1990th (Smith and Rayment, 1996; Kinoshita et al., 1999), introducing a whole universe of the P-loop constellations known today (Goncearenco and Berezovsky, 2010, 2015; Berezovsky, 2019; Longo et al., 2020a; Zheng et al., 2016). For example, some works focused on the recognition and binding of phosphate in specific ligands or groups thereof. For example, importance of the phosphate binding signature for molecular recognition was shown in the analysis of four different FAD-binding protein families: glutathione reductase, ferredoxin reductase, p-cresol methyl hydroxylase, and pyruvate oxidase (Dym and Eisenberg, 2001). The Adenine recognition in ATP, CoA, NAD, NADP, FAD and other adenine -containing ligands was shown to be provided by the "adenine-binding motif present in ancient proteins and common to all current structures" (Denessiouk et al., 2001). With an increase of the sequence/structure data it became apparent that few major structural types, including P-loop, FAD/NAD(P)-binding fold, Rossmann-like folds, represent the diversity of the phosphate binding in different folds and functions (Brakoulias and Jackson, 2004). The structure-based derivation of the phosphate binding sequence motif revealing major sequence signatures was shown to be an alternative way for obtaining the typical signatures on the basis of the 3D conservation of corresponding structural motifs (Hua et al., 2014).

Several works produced by the Dan Tawfik's group and his collaborators focused on the evolutionary aspects of the phosphate binding, from the very beginning in prebiotic evolution to diversification in different biochemical functions and its subsequent evolution withing corresponding folds and their enzymatic activities. For example, exploring Rossmann and Rossmann-like structures that bind different nucleotide-containing cofactors authors showed that they served as a platform for distinct chemistries taking place in corresponding biochemical functions (Laurino et al., 2016). Further studies also showed that many domains that bind (di)nucleotide-containing signatures, such as HUP, flavodoxin, TIM-barrel, P-loop, and different Rossmanns emerged from the short peptide containing the phosphate binding signature currently located in the N-helix site (Longo et al., 2020b). It was also shown that simple P-loop signature in β-turn-α polypeptides possess weak helicase function, being capable to bind ssDNA and RNA and to facilitate unwinding dsDNA upon addition of NTPs or inorganic phosphates (Vyas et al., 2021). The work on HUP domain showed an interesting example of the phosphate bindings' evolutionary usage: it was complemented by another conserved signature of the ribose binding, resulting in a specification by addition of new elementary function (Gruic-Sovulj et al., 2022). One more example of the function built around the phosphate binding is the CoA-binding Nat/Ivy protein (Longo et al., 2022b). This study was inspired by late Dan Tawfik, showing his great legacy in the protein function and evolution research in general and illuminating contribution into the understanding of one of the key elementary functions – the phosphate binding (Romero Romero and Rabin, 2016; Laurino et al., 2016; Longo et al., 2020a, 2020b; Romero et al., 2018).

The fundamental Tawfik's group contribution in the above topic is summarized by the work on P-loop NTPases and Rossman proteins (Longo et al., 2020a). It shows that while the former catalyzes the phosphoryl transfer, the latter provides a handle for binding of different ligands, raising a question a question about divergence versus convergence as an evolutionary scenario of their emergence. At the same time, the work clearly demonstrates that all considered cases are based on the common β-α-β motif (also described as a closed loop (Berezovsky et al., 2000; Berezovsky and Trifonov, 2001a; Berezovsky and Trifonov, 2001b; Berezovsky et al., 2002) or supersecondary structure element (Levitt and Chothia, 1976)) dominated by Gly (Trifonov, 1999) and few other ancient amino acids (Trifonov, 2000), allowing one to hypothesize the common origin of *all* phosphate binding elementary functions in the whole diversity of contemporary proteins. One of the most recent works inspired by Dan Tawfik and finished by his colleagues further supports this hypothesis, unraveling the key role of the phosphoryl transfer as a fundamental reaction in the cellular metabolism existing from the origin of enzymes (Vyas et al., 2023). The whole spectrum of elementary functions with a detail description of the biding specifics of all components of (di)nucleotide-containing ligands is accumulated in the nucleotide binding database (NBDB, https://nbdb.bii.a-star.edu.sg (Zheng et al., 2016),), which became an inspiration and reliable data source for a number of works exploring the evolution, function, and design potential of the (di)nucleotide-containing ligand binding (Heizinger and Merkl, 2021; Nepomnyachiy et al., 2017; Yin et al., 2021; Longo et al., 2020a, 2020b, 2022b; Chu and Zhang, 2020; Kolodny, 2021; Romero-Romero et al., 2021; Narunsky et al., 2020; Bhagavat et al., 2017).

2.4. The phosphate binding and derivation of its ancient prototype

Diversity of the GxGxxG's representatives of the phosphate binding in dinucleotide-containing ligands discussed above (Fig. 1), its high similarity to the phosphate binding in nucleotide-containing ligands (with GxxGxG generic signature), as well as earlier observations hinting on their possible common origin from more generalized and simplified Gly-rich prototype (Goncearenco and Berezovsky, 2010, 2015; Berezovsky, 2019) call for a *back in time* journey to see potentially most ancient simple prototype of the phosphate binding. It would be important to reconstruct it, to see peculiarities that made it work then, in the very beginning of protein evolution, and provided a potential for evolving into a current repertoire of the structure- and functional (super) family-specific corresponding elementary functions (Goncearenco and Berezovsky, 2010, 2015; Berezovsky, 2019). To this end, we performed here a simple experiment on the prototype reconstruction, which is briefly described and illustrated below. The phosphate binding profiles were derived based on 10,804 structures downloaded from the Protein Data Bank (PDB), which contain 23 ligands (Suppl. Table S1). The 30-residue long sequences comprise the dataset of phosphate binding fragments collected from the above PDB structures. These fragments were originally grouped based on the similarity in protein-ligand interactions described in corresponding non-redundant PDB structures. The set of profile-origins (or origins) in form of PSSM, where each origin was built from corresponding group of sequences, was obtained (See Appendix in Suppl. File 1 and Suppl. Figures 1and 2). The glycine-enriched signatures presented in Fig. S3 were annotated and subjected to iterative merging procedure (see Appendix in Suppl. File 1 for details). The goal was to obtain the generalized prototype of the phosphate binding signatures in (di)nucleotide-containing ligands. The profile signature-detection power is increasing upon merger (Suppl. Fig. S4), yielding more generic profiles that recognize more individual signatures of different mono- and dinucleotide containing ligands. The procedure based on the profiles' similarity (Suppl. Fig. S3) resulted in the glycine-rich prototype with GxGGxG characteristic signature (Fig. 2).

Fig. 2 presents the "Phosphate-Binding Prototype Circle" – a diagram, illustrating a relationship between the sequences and structures of MONO/DI profiles and of the most generic Gly-rich PROTOTYPE obtained in the reconstruction procedure (See Appendix and Suppl. Figs. S1–S3). The diagram also shows few typical specific representatives working in the phosphate binding of different (di)nucleotide-

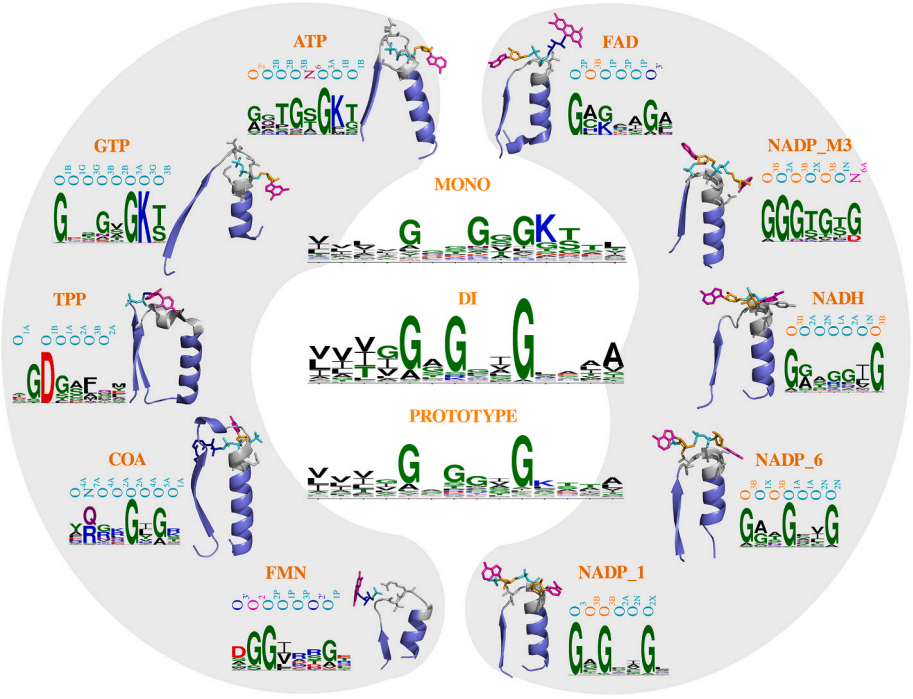


Fig. 2. The "Phosphate-Binding Prototype Circle" – a diagram of the relationship between the DI-/MONO-Profiles and the Gly-rich Prototype. DI and MONO signatures in the center show the generalized profiles for the phosphate binding in nucleotide- and dinucleotide-containing ligands. The PROTOTYPE logo describes presumable ancient ancestor of the glycine-rich phosphate-binding signatures that exist in contemporary proteins.

containing ligands. The MONO profile is built from signatures of profiles that recognize the nucleotide-containing ligands, whereas DI profile – from the sequences of elementary functional loops binding the dinucleotide-containing ligands. Notably, the representative signatures of the MONO profile mostly shows interactions with phosphate group regardless of the type of the ligand. The signatures of the DI profile reveal more diverse interactions with several moieties of corresponding ligands.

Fig. 3 presents selected examples of interactions between the (di) nucleotide-containing ligands and the phosphate-binding signatures, illustrating characteristic distinctions in corresponding sequence patterns of the phosphate binding. Specifically, most of the patterns for dinucleotide-containing ligands reveal three positions with high frequency of glycine in them contained in the six-seven residue central segment, which is flanked by four-residue hydrophobic pattern (with preference for valine, isoleucine, and leucine in order of decreasing frequency of these residues) on the left side and notably preferred alanine on the fourth position of the right flank. The signatures of phosphate binding in nucleotide-containing ligands are characterized by the preference of glycine in more positions, which is also complemented by the presence of charged residues (aspartic/glutamic acids and lysine) in some other positions. Noteworthy, higher presence of glycine in several positions in case of interactions with nucleotide-containing ligands coincides with domination of signature's interactions with phosphate groups of ligands, including nucleotide-free PLP and TPP, which may serve as another indication of the phosphate binding as the original function of the Gly-rich signature.

Fig. 4 provides further details of the most conserved interactions typical for the generic Gly-rich prototype (center) in case of nucleotide- (center, top) and dinucleotide-containing (center, bottom) ligand binding. The most conserved interactions detected by the prototype are shown in the central row: for nucleotide-containing (upper histogram) and dinucleotide-containing (bottom histogram) ligands. Conservation of interactions for the MONO (top) and DI (bottom) profiles and for their contributions to PROTOTYPE (middle) is calculated as frequencies of interactions between residues of binding signatures comprising the profile and atoms of bound ligands. The highest frequency contacts are shown in corresponding histograms in Fig. 4. Conservation of contacts observed with generalized profiles of the phosphate binding in nucleotide-containing (top level) and dinucleotide-containing (bottom level) ligands is in a good agreement with those obtained for the PROTOTYPE (central level). It also shows more specific conserved interactions observed for MONO and DI profiles representing the phosphate binding in nucleotide- (top) and dinucleotide-containing (bottom). Despite less diversity of dinucleotide-containing ligands, the DI profiles and PROTOTYPE's signatures show interactions with all three moieties of ligands, sugar, phosphate, and base, which are determined mostly by the interactions with NAD. The MONO profiles reveal only two positions on the left interaction with ribose, while others make contacts with the phosphate. Noteworthy, while conserved interactions of the MONO profile are chiefly determined by Adenine/Guanine-containing ligands, the Gly-rich prototype's interactions are strongly affected by the PLP-protein interactions (Fig. 3). In case of DI profile, its interactions (histogram in the bottom) are very similar to those of the Prototype (bottom histogram in the center).

Fig. 5 shows the connection between the binding specificity of signatures, profiles, and the Prototype in relation to similarity between bound ligands. The ligand similarity consideration clearly detects two groups of the most present nucleotide-containing (ATP, ADP, GDP) dinucleotide-containing (NAD(H/P), FAD) ligands. Of note, nucleotide-containing ligands are well recognized by both generalized profiles of the phosphate binding in nucleotide-containing (Mono_Profiles) and dinucleotide-containing (Di_Profiles) ligands, as well as by the very general Prototype, which was the goal of this reconstruction. The dinucleotide-containing group (NAD(H/P), FAD), however, is preferentially recognized by its specific profiles and by the Prototype. Interestingly, that one of the simplest ligands not even having the nucleotide moiety at all, but included as one of the most relevant simple phosphate-containing ligands, the Pyridoxal-5'-phosphate (PLP), is also recognized by only Di_Profile and its more specific representatives. These observations agree with previously suggested role of the phosphate binding in dinucleotide-containing ligands as a handle for binding of other

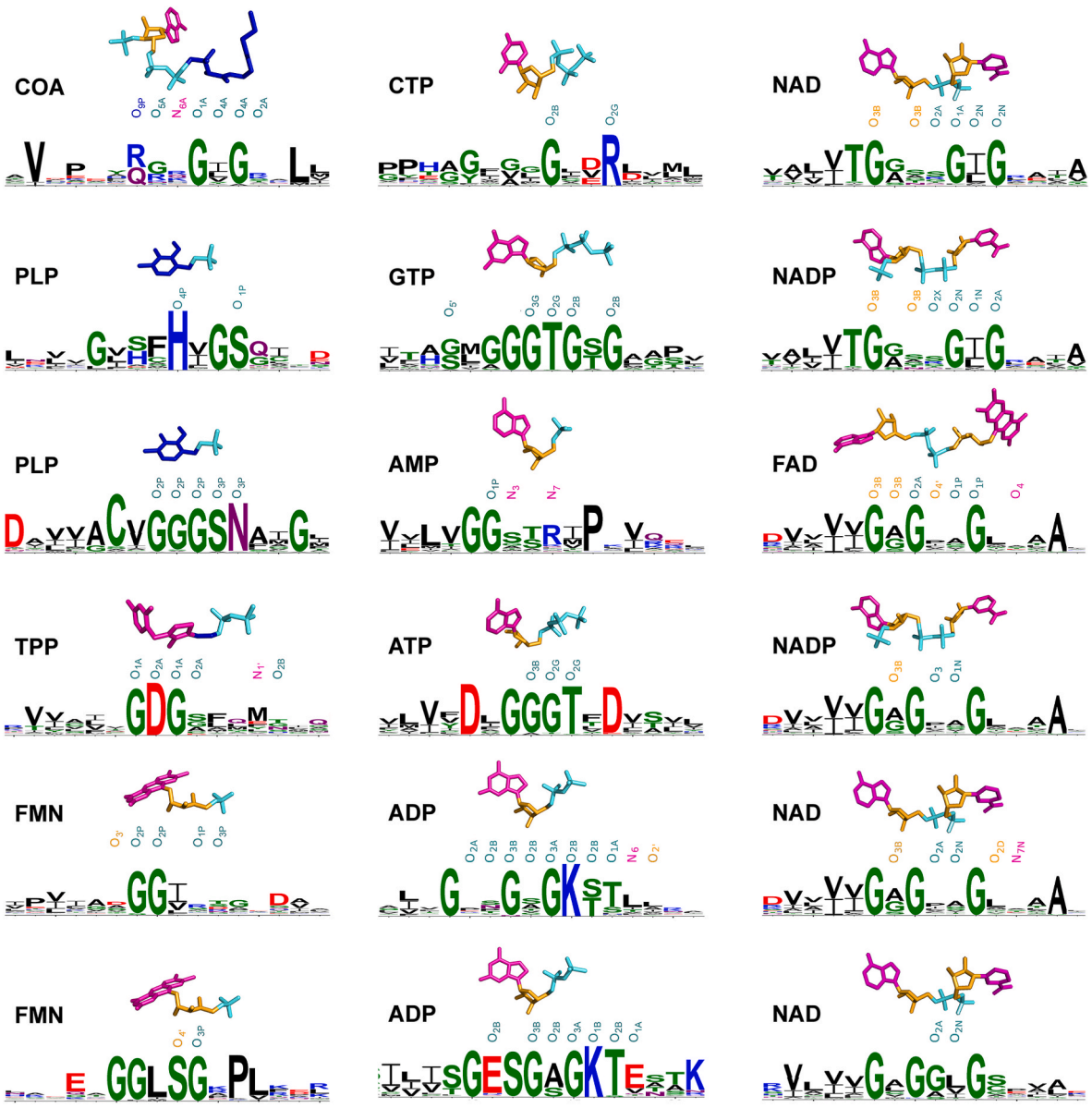


Fig. 3. Examples of G-enriched profiles in the set of considered signatures. Examples of atom interactions via hydrogen bonds between (di)nucleotide-containing ligands and signatures. The binding of ligands of mono-nucleotide specific signatures extensively relies on the hydrogen bonds between phosphate moiety and glycines. The dinucleotide binding signatures interact with ligands in a more diverse way, including more interactions with other ligand moieties, especially ribose.

substrates (so-called, Rossmann signatures with generalized pattern GxGxxG), while the phosphate binding in nucleotide containing ligands (P-loop NTPases, GxxGxGK[TS]) are considered as functional motifs facilitating the phosphoryl transfer (Noor et al., 2022). More details on conservatism of interactions between the phosphate-binding signatures and the atoms of ligands mapped on the DI-/MONO-profiles and on the Gly-rich Prototype are provided in Suppl. Fig. S5.

Lastly, we review a structural diversity of folds with functions that involve the binding of (di)nucleotide-containing ligands. The structural annotation of proteins containing specific and generalized profiles (Mono_profile, Di_profile, and the Prototype shown as GxGGxG) reveals domination of the α/β and α+β folds (Fig. 6). The α/β-sandwiches (typically, three-layer, including c.37.1, c.36.1, c.48.1 c.55.1, and others) and β/α-barrels (*e.g.,* c.1.4 and c.1.5) constitute most fold types representing the α/β (c-class) proteins. The major basic unit of both types of above folds is the β-α-β closed loop (Berezovsky and Trifonov, 2001a, 2001b, 2002a) or its variant – the β-α return of the protein backbone – built in and stabilized in the context of the overall fold (Goncearenco and Berezovsky, 2010, 2015; Berezovsky et al., 2017b). In both cases the underlying structural unit is apparently a descendant of the ancient ring-like peptide, a potential member of the "Dayhoff fragment" cohort, which is embedded in modern folds (Berezovsky et al., 2000, 2002, 2017b; Berezovsky and Trifonov, 2001a, 2002a). The generic polymer origin of closed loops regardless of their secondary structure compositions or those of the overall folds is further corroborated by the analysis of the α+β (d-class) proteins (*e.g.,* d.48.1, d.56.1, d.58.2, d.128.1 and others in Fig. 6). These folds are also characterized by mostly layered architectures formed from an extended repertoire (complemented, for example, by the α- and β-hairpins) of protein chain returns. Thus, a persistence of protein chain returns (Berezovsky et al., 2002) and its omnipresence in all protein classes supports the emergence of modern domains/folds as combinations of "Dayhoff fragments" in form of ring-like peptides (Goncearenco and Berezovsky, 2012, 2015; Berezovsky et al., 2017b), which were eventually decorated with the secondary structure elements (Berezovsky and Trifonov, 2001a, 2002a) that facilitated diversification into different architectures emerged in the

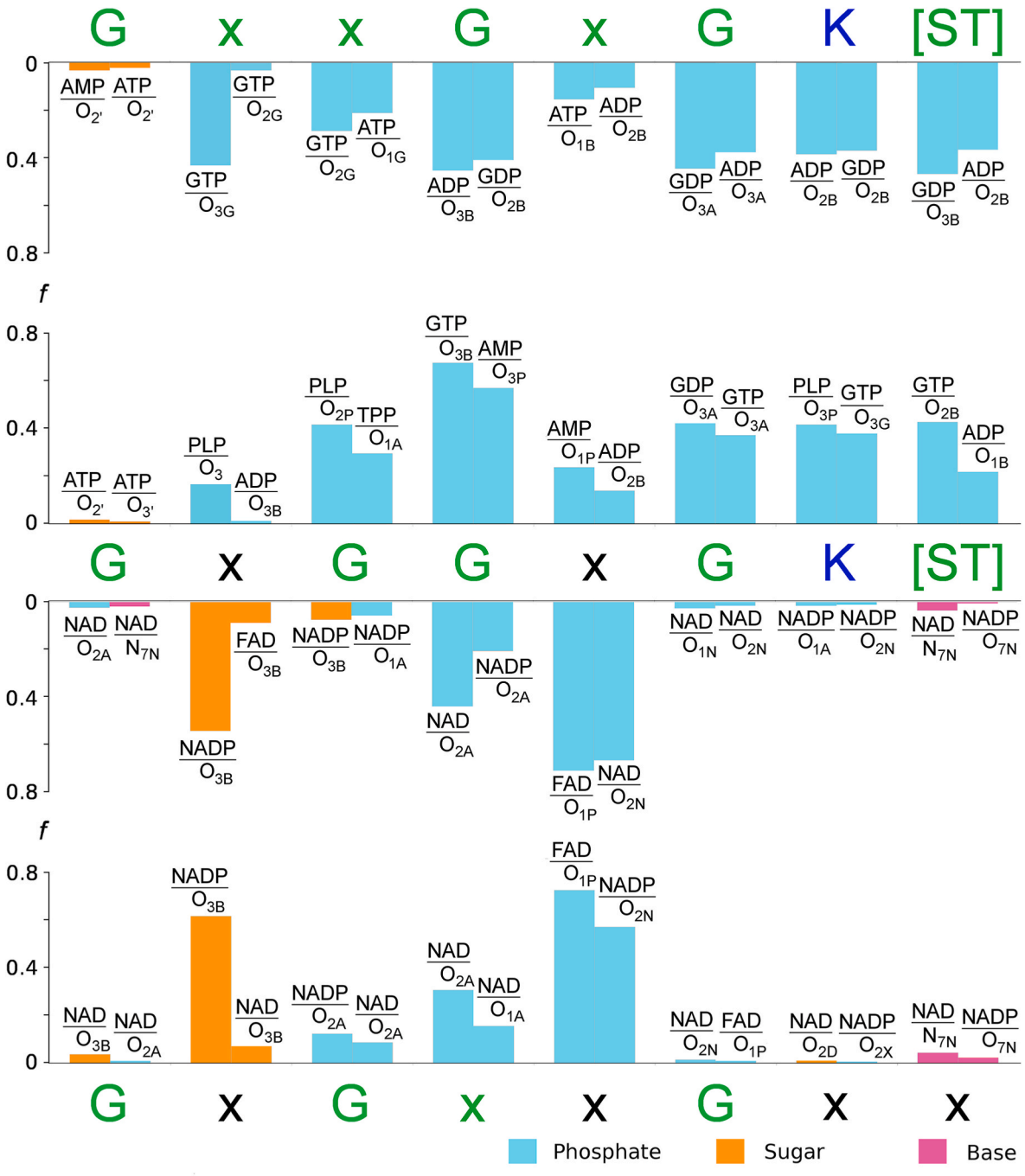


Fig. 4. The most conserved interactions between residues of the phosphate binding signatures and atoms of the (di)nucleotide-containing ligands.

evolution of protein structure and function. Importantly, the phylogenomic analysis of gene ontology data (Koc and Caetano-Anolles, 2017) complements a picture of the evolution of molecular function with a chronology of cofactors' usage by SCOP families (Murzin et al., 1995).

Two major superfamilies are p-loop NTPasas detected by the Mono_profile and NAD(P)-binding Rossmann fold domains found by the Di_profile. In agreement with an observation on the similarity of profiles and their potential to find matches on corresponding proteins, Di-profile can detect elementary functions of the phosphate binding also in nucleotide-containing ligands present in P-loop NTPases (c.37.1, c.111.1, and some others). Notably, the Di_profile finds elementary function of the phosphate binding for ATP and ADP ligands and, in general, it seems to be very generic detecting the binding in most of the functional families. It can also be related to the presence of only three major profiles working in dinucleotide-containing ligand binding, NAD(H), NADP, and FAD(H), whereas nucleotide-containing ligands require many distinct and more specific profile signatures. (Fig. 6). It also agrees with the more demanding role of the P-loop working on the phosphoryl transfer, helicase activity, and its involvement into binding of different cofactors all of which require more specificity of sequence and structure. The Di_profile characteristic for Rossmann and Rossmann-like folds indicate the function of the "handle provider for binding of other substrates", hence less specific and working generically in different biochemical functions (Noor et al., 2022). Comparative analysis of proteomes corroborates existence of common and widely reused structural, functional, end evolutionary units (Caetano-Anolles et al., 2021). Starting from building folds from the Dayhoffs ring-like peptides, descendants of which were described as super-secondary structure

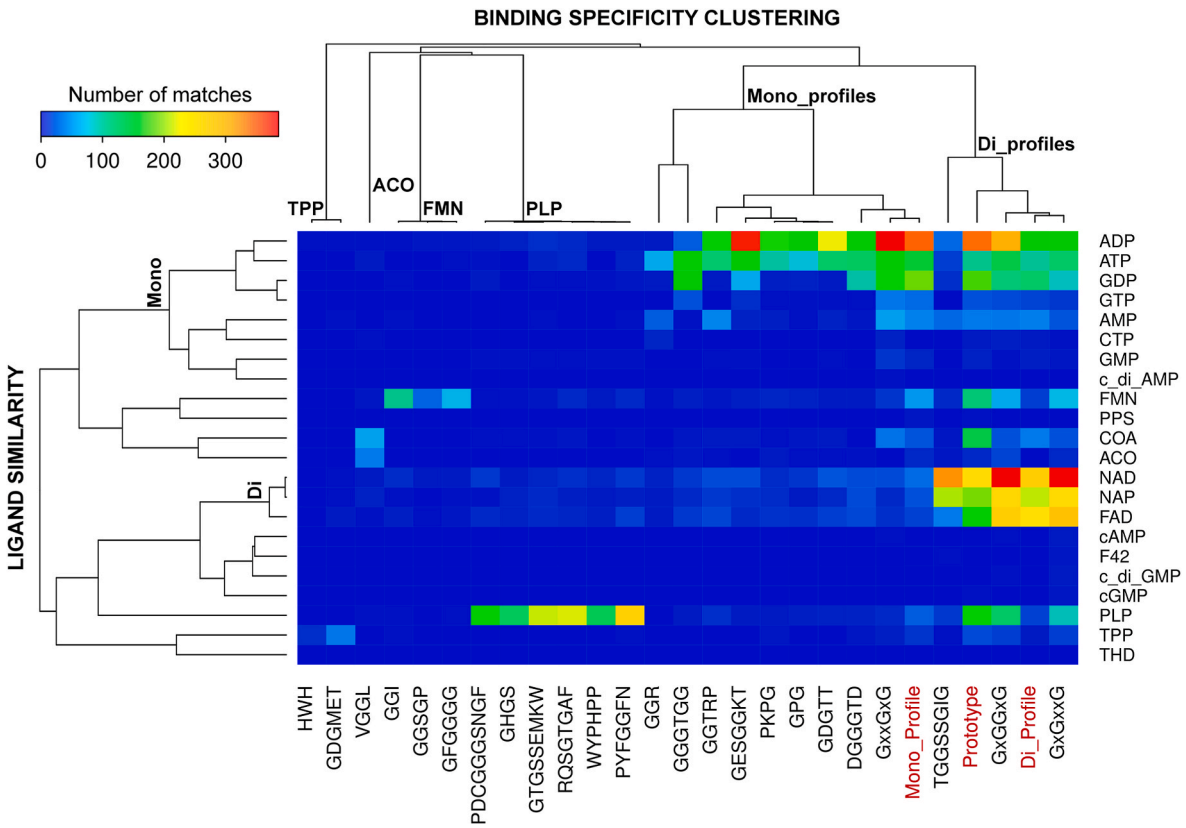


Fig. 5. Ligand similarity and binding specificity-based analysis of the profile similarity. The "ligand vectors" (rows) are clustered based on similarity of ligands' binding modes manifested in corresponding profiles. The "profile vectors" (columns) are clustered based on the profiles "binding" signatures' detection power. The comparison corroborates that signatures of the phosphate binding in dinucleotide-containing ligand binding are more generic than those that work in nucleotide-containing ligand binding.

elements by Levitt and Chothia (1976) and presented in modern proteins in form of the returns of the protein backbones (Berezovsky et al., 2000, 2002; Berezovsky and Trifonov, 2001a, 2001c), the evolution proceeded into building of new and more complex structures, gaining new functions, and their combinations (Caetano-Anolles et al., 2021). A graph-theoretical approach aimed at tracing the emergence of protein domains from loops revealed functional links between repertoires of loop prototypes and folded structural domains (Aziz et al., 2023). The chronologies of domain structures and architectures unravel, in turn, emergence of major fold types, such as barrels, sandwiches, and bundles, followed by their involvement in higher order assemblies and combinations coincided with further evolution and diversification of their functions (Caetano-Anolles et al., 2021).

3. Conclusions

An understanding of the protein function that would allow one to perform *de novo* design should be based on the knowledge of its basic elementary units with structural or catalytic roles that they perform and, desirably, on the understanding of physical and evolutionary mechanisms of their evolutionary persistence. Only several dozens of distinct chemical roles of catalytic residues, making about 400 functional mechanism that underlie some 5000 currently know biochemical transformations, call for the detail study of corresponding units of proteins (Goncearenco and Berezovsky, 2015). Based on a common agreement that interactions between the nucleic acids and proteins were present in the very beginning and had facilitated Origin of Life, we reviewed and analyzed here one of the most common and ancient functions – the phosphate binding. Already early works on a very limited data showed diversity of roles (Goncearenco and Berezovsky, 2010, 2015; Longo et al., 2020a, 2020b) playing by the phosphate binding and by the relevant ligands/co-factors (Dym and Eisenberg, 2001; Denessiouk et al., 2001). The common origin of this elementary function was discussed from the perspective of both structure and function, pointing to β-α-β closed loop (Berezovsky et al., 2000, 2017b; Berezovsky and Trifonov, 2001a) as a potential first structural carrier of the short phosphate binding signature (Goncearenco and Berezovsky, 2010, 2015; Longo et al., 2020b), which is still detectable in modern proteins (Berezovsky, 2019; Goncearenco and Berezovsky, 2012; Schneider et al., 2006; Laurino et al., 2016; Longo et al., 2020a). It was also shown that phosphate binding became an element of diverse biochemical function, contributing weak helicase activity (Vyas et al., 2021), providing transfer of the phosphoryl group (Vyas et al., 2023), and being complemented by other elementary functions (Berezovsky et al., 2003a, 2003b; Gruic-Sovulj et al., 2022; Longo et al., 2022b). Based on numerous works illuminating the sequence/structure determinants, enzymatic and other biochemical reactions it is part of, we reconstructed the most ancient, simplest, and versatile prototype from which this elementary function apparently started.

The biophysical perspective on the emergence and early evolution of protein structure and function supports the model of building complex multistep biochemical transformation from simple reactions provided by elementary units (Goncearenco and Berezovsky, 2015; Berezovsky et al., 2000, 2017b). It proposes the loop closure based on the polymer nature of protein polypeptide chains as an advantageous event in prebiotic evolution, which resulted in the emergence of the first ring-like peptides with elementary functions (Goncearenco and Berezovsky, 2015; Romero Romero and Rabin, 2016; Eck and Dayhoff, 1966; Berezovsky et al., 2017b). They later formed first functional protein folds (Goncearenco and Berezovsky, 2015; Berezovsky et al., 2017b) thanks to fusion of respective small genes (Romero Romero and Rabin, 2016; Eck and Dayhoff, 1966; Trifonov and Berezovsky, 2003). The most common

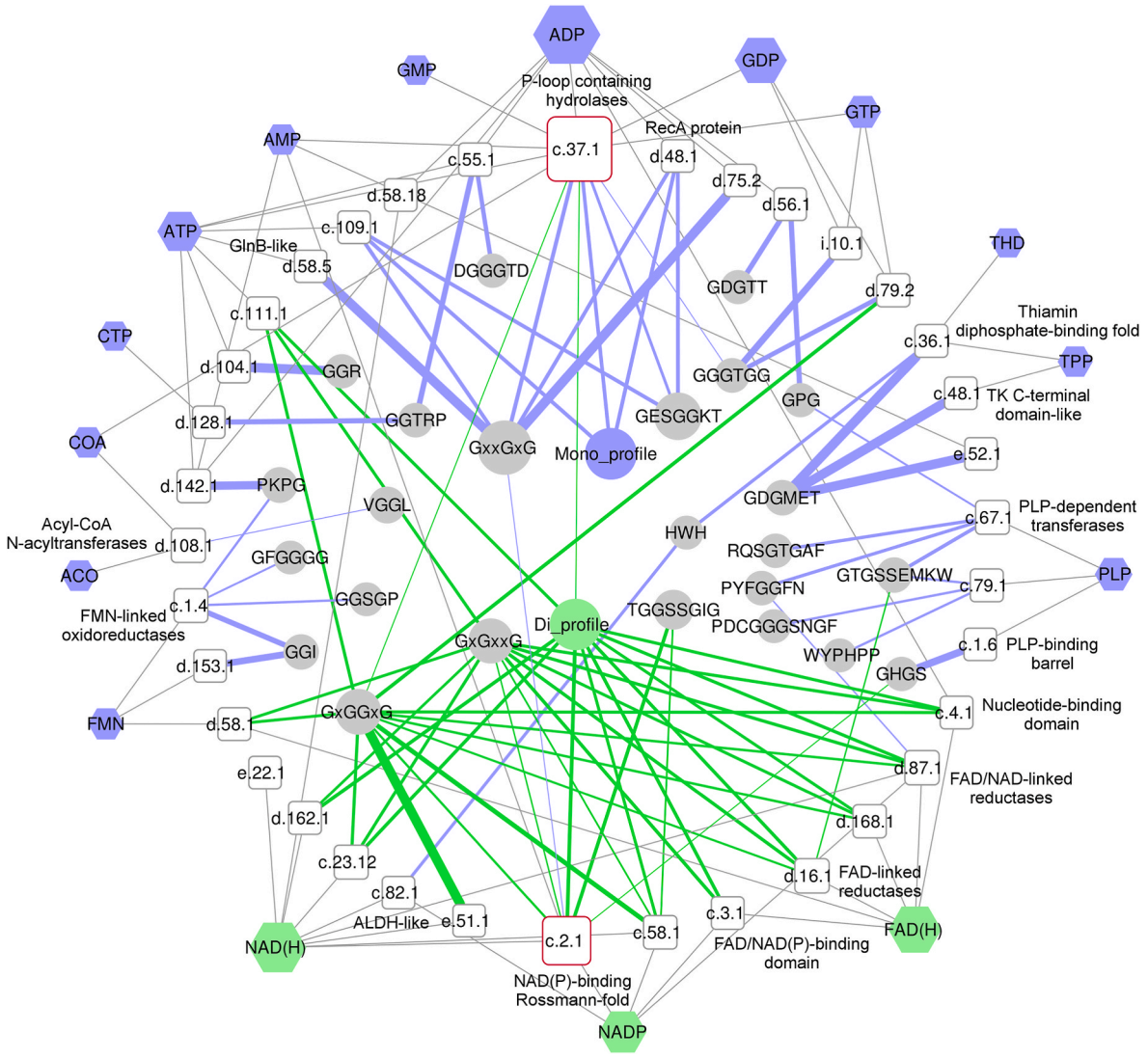


Fig. 6. Structural annotation of profiles in SCOP superfamilies. The matches between profile PSSMs and sequences of SCOPe superfamilies show representation of profiles in different folds and function. Glycine-rich signatures of the phosphate-binding in (di)nucleotide-containing ligands and their matches to sequences of proteins of SCOP superfamilies. The phosphate binding signatures (GxGxxG, green) of dinucleotide-containing ligands connects with a wide spectrum of SCOP superfamilies, while those of nucleotide-containing ligands (GxxGxG, blue) show limited connections to SCOP superfamilies mostly specific its own (super)family.

β-α-β structural unit discovered almost 50 years ago (Levitt and Chothia, 1976) and properly appreciated in recent works as the major element carrying the phosphate binding in modern proteins (Laurino et al., 2016; Longo et al., 2020a) resembles the descendant of one of the first prebiotic functional ring-like peptides (Goncearenco and Berezovsky, 2015; Berezovsky et al., 2000, 2017b).

How did these first primitive enzymes look like? What were their sequences that gave rise to the current repertoire, providing the specificity and diversity of structures and functions? In attempt to answer these questions, we performed here a simple computational experiment, reconstructing the most basic, ancient prototype of the phosphate binding elementary function. Resulting prototype appeared to be Gly-rich in agreement with earlier conclusions on the origin and evolution of the genetic code and the temporal order of amino acids (Trifonov, 1999, 2000). Several other, non-glycine amino acids presented in the prototype, such as alanine and valine, occupying many positions of the prototype are also ancient ones according to the amino acid chronology (Trifonov et al., 2001; Trifonov, 2000). This simplified signature determined a high versatility of the evolutionary design, allowing it to perform simple functions (*e.g.* helicase (Vyas et al., 2021) or phosphoryl transfer (Vyas et al., 2023)) or binding of different ligands, cofactors, or providing a handle for the other substrate binding (Goncearenco and Berezovsky, 2010, 2015; Laurino et al., 2016; Longo et al., 2020a). Such function-binding duality possessed by presumed ancient functional peptides agrees with one of the most recent ideas of uniform binding proposed by Dan Tawfik. The latter shows that presence of "chemistry-first" or "substrate-binding-first" in the beginning of protein evolution provided a critical versatility of the evolutionary design and resulted in a diversity of contemporary protein functions (Noor et al., 2022). It remains to be seen how this exciting idea will be further supported, developed, and used, enriching the modern engineering and design of new proteins (Vyas et al., 2021, 2023; Romero et al., 2018).

Finally, one may ask a question why to worry about so distant past in the times of the AlphaFold (Varadi et al., 2022), trying to go back four billion years to only find a very simple, rather primitive peptides performing very basic functions? The answer is that for millions of proteins with fine-tuned functions specific to different organisms and their life styles, there are only about 5000 biochemical transformations (Bairoch, 2000) provided by about 400 unique mechanism (Holliday et al., 2012), which, in turn, determined by about 30 basic chemical roles of catalytic amino acids (Holliday et al., 2005; Andreini et al., 2009). Therefore, starting from consideration of a very limited repertoire of basic

elementary functions, using the most generic, hence versatile for future design and tuning, sequences and structures possessing these functions, the whole diversity of contemporary biological functions can be obtained. The repertoire of evolutionary selected and weathered diversity of functions can be used together with and complemented by the AlphaFold predictions may strongly facilitates rational design of desirable protein functions. The "back in time" travel here showed only one, but excellent example of simple elementary function arrived from a prebiotic world and turned into a cornerstone of multiple protein functions. It shows a great importance of the early protein evolution studies, which can facilitate not only engineering and *de novo* design of functions, but can also be instructive in introducing not yet fully appreciated and used mechanisms of their regulation (Guarnera and Berezovsky, 2016, 2019a, 2019b; Tee et al., 2021, 2022; Berezovsky and Nussinov, 2022).

Credit author statement

ZZ performed the work, analyzed data; AG analyzed data, INB supervised the work, analyzed data, wrote and edited paper.

Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported by the core funding provided by the Biomedical Research Council (BMRC) of the Agency for Science, Technology, and Research (A*STAR), Singapore. INB was also partially supported by the NMRC MOH-001402-00 grant. AG and INB dedicate this work to the memory of Professor Dan (Danny) Tawfik - dear friend and colleague.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crstbi.2024.100142.

The interaction frequencies are represented by the heights of the bars and the ligand moieties are coded by the colours (blue – phosphate; orange - sugar; magenta - base). Most of the positions of the nucleotide-binding profile are involved in conserved interactions with phosphate groups, only first glycine interacts with the sugar. The histograms show the highest frequency contacts observed between residues of signatures and atoms of bound ligands. In case of the dinucleotide-containing ligands, only two central positions are in conserved interactions with the phosphate group, whereas first two positions interact with sugar and last two positions – with base.

References

Alva, V., Soding, J., Lupas, A.N., 2015. A vocabulary of ancient peptides at the origin of folded proteins. Elife 4, e09410.
Amangeldina, A., Tan, Z.W., Berezovsky, I.N., 2024. Living in trinity of extremes: Genomic and proteomic signatures of halophilic, thermophilic, and pH adaptation. Curr Res Struct Biol 7, 100129.
Andreini, C., Bertini, I., Cavallaro, G., Holliday, G.L., Thornton, J.M., 2009. MetalMACiE: a database of metals involved in biological catalysis. Bioinformatics 25, 2088–2089.
Angelov, B., Sadoc, J.F., Jullien, R., Soyer, A., Mornon, J.P., Chomilier, J., 2002. Nonatomic solvent-driven Voronoi tessellation of proteins: an open tool to analyze protein folds. Proteins 49, 446–456.
Aziz, M.F., Caetano-Anolles, G., 2021. Evolution of networks of protein domain organization. Sci. Rep. 11, 12075.
Aziz, M.F., Caetano-Anolles, K., Caetano-Anolles, G., 2016. The early history and emergence of molecular functions and modular scale-free network behavior. Sci. Rep. 6, 25058.
Aziz, M.F., Mughal, F., Caetano-Anolles, G., 2023. Tracing the birth of structural domains from loops during protein evolution. Sci. Rep. 13, 14688.
Bairoch, A., 2000. The ENZYME database in 2000. Nucleic Acids Res. 28, 304–305.
Bella, J., Hindle, K.L., McEwan, P.A., Lovell, S.C., 2008. The leucine-rich repeat structure. Cell. Mol. Life Sci. 65, 2307–2333.
Ben Ishay, E., Rahamim, G., Orevi, T., Hazan, G., Amir, D., 2012. Haas E: fast subdomain folding prior to the global refolding transition of E. coli adenylate kinase: a double kinetics study. J. Mol. Biol. 423, 613–623.
Berezovsky, I.N., 2003. Discrete structure of van der Waals domains in globular proteins. Protein Eng. 16, 161–167.
Berezovsky, I.N., 2011. The diversity of physical forces and mechanisms in intermolecular interactions. Phys. Biol. 8, 035002.
Berezovsky, I.N., 2019. Towards descriptor of elementary functions for protein design. Curr. Opin. Struct. Biol. 58, 159–165.
Berezovsky, I.N., Nussinov, R., 2022. Multiscale allostery: basic mechanisms and versatility in diagnostics and drug design. J. Mol. Biol. 434, 167751.
Berezovsky, I.N., Shakhnovich, E.I., 2005. Physics and evolution of thermophilic adaptation. Proc. Natl. Acad. Sci. U. S. A. 102, 12742–12747.
Berezovsky, I.N., Trifonov, E.N., 2001a. Loop fold nature of globular proteins. Protein Eng. 14, 403–407.
Berezovsky, I.N., Trifonov, E.N., 2001b. Van der Waals locks: loop-n-lock structure of globular proteins. J. Mol. Biol. 307, 1419–1426.
Berezovsky, I.N., Trifonov, E.N., 2001c. Protein structure and folding: a new start. J. Biomol. Struct. Dyn. 19, 397–403.
Berezovsky, I.N., Trifonov, E.N., 2002a. Flowering buds of globular proteins: transpiring simplicity of protein organization. Comp. Funct. Genom. 3, 525–534.
Berezovsky, I.N., Trifonov, E.N., 2002b. Loop fold structure of proteins: resolution of Levinthals paradox. J. Biomol. Struct. Dyn. 20, 5–6.
Berezovsky, I.N., Tumanyan, V.G., Esipova, N.G., 1997. Representation of amino acid sequences in terms of interaction energy in protein globules. FEBS Lett. 418, 43–46.
Berezovsky, I.N., Namiot, V.A., Tumanyan, V.G., Esipova, N.G., 1999. Hierarchy of the interaction energy distribution in the spatial structure of globular proteins and the problem of domain definition. J. Biomol. Struct. Dyn. 17, 133–155.
Berezovsky, I.N., Grosberg, A.Y., Trifonov, E.N., 2000. Closed loops of nearly standard size: common basic element of protein structure. FEBS Lett. 466, 283–286.
Berezovsky, I.N., Kirzhner, V.M., Kirzhner, A., Trifonov, E.N., 2001. Protein folding: looping from hydrophobic nuclei. Proteins 45, 346–350.
Berezovsky, I.N., Kirzhner, V.M., Kirzhner, A., Rosenfeld, V.R., Trifonov, E.N., 2002. Closed loops: persistence of the protein chain returns. Protein Eng. 15, 955–957.
Berezovsky, I.N., Kirzhner, A., Kirzhner, V.M., Rosenfeld, V.R., Trifonov, E.N., 2003a. Protein sequences yield a proteomic code. J. Biomol. Struct. Dyn. 21, 317–325.
Berezovsky, I.N., Kirzhner, A., Kirzhner, V.M., Trifonov, E.N., 2003b. Spelling protein structure. J. Biomol. Struct. Dyn. 21, 327–339.
Berezovsky, I.N., Guarnera, E., Zheng, Z., Eisenhaber, B., Eisenhaber, F., 2017a. Protein function machinery: from basic structural units to modulation of activity. Curr. Opin. Struct. Biol. 42, 67–74.
Berezovsky, I.N., Guarnera, E., Zheng, Z., 2017b. Basic units of protein structure, folding, and function. Prog. Biophys. Mol. Biol. 128, 85–99.
Bergasa-Caceres, F., Rabitz, H.A., 2018. Predicting the location of the non-local contacts in alpha-synuclein. Biochim. Biophys. Acta, Proteins Proteomics 1866, 1201–1208.
Bhagavat, R., Srinivasan, N., Chandra, N., 2017. Deciphering common recognition principles of nucleoside mono/di and tri-phosphates binding in diverse proteins via structural matching of their binding sites. Proteins 85, 1699–1712.
Bharat, T.A., Eisenbeis, S., Zeth, K., Hocker, B., 2008. A beta alpha-barrel built by the combination of fragments from different folds. Proc. Natl. Acad. Sci. U. S. A. 105, 9942–9947.
Blaber, M., Lee, J., 2012. Designing proteins from simple motifs: opportunities in Top-Down Symmetric Deconstruction. Curr. Opin. Struct. Biol. 22, 442–450.
Brakoulias, A., Jackson, R.M., 2004. Towards a structural classification of phosphate binding sites in protein-nucleotide complexes: an automated all-against-all structural comparison using geometric matching. Proteins 56, 250–260.

Bromberg, Y., Aptekmann, A.A., Mahlich, Y., Cook, L., Senn, S., Miller, M., Nanda, V., Ferreiro, D.U., Falkowski, P.G., 2022. Quantifying structural relationships of metal-binding sites suggests origins of biological electron transfer. Sci. Adv. 8, eabj3984.
Broom, A., Doxey, A.C., Lobsanov, Y.D., Berthin, L.G., Rose, D.R., Howell, P.L., McConkey, B.J., Meiering, E.M., 2012. Modular evolution and the origins of symmetry: reconstruction of a three-fold symmetric globular protein. Structure 20, 161–171.
Brunette, T.J., Parmeggiani, F., Huang, P.S., Bhabha, G., Ekiert, D.C., Tsutakawa, S.E., Hura, G.L., Tainer, J.A., Baker, D., 2015. Exploring the repeat protein universe through computational protein design. Nature 528, 580–584.
Caetano-Anolles, G., Aziz, M.F., Mughal, F., Caetano-Anolles, D., 2021. Tracing protein and proteome history with chronologies and networks: folding recapitulates evolution. Expert Rev. Proteomics 18, 863–880.
Chen, C.K., Chan, N.L., Wang, A.H., 2011. The many blades of the β-propeller proteins: conserved but versatile. Trends Biochem. Sci. 36, 553–561.
Chintapalli, S.V., Illingworth, C.J., Upton, G.J., Sacquin-Mora, S., Reeves, P.J., Mohammedali, H.S., Reynolds, C.A., 2014. Assessing the effect of dynamics on the closed-loop protein-folding hypothesis. J. R. Soc. Interface 11, 20130935.
Chu, X.Y., Zhang, H.Y., 2020. Cofactors as molecular fossils to trace the origin and evolution of proteins. Chembiochem 21, 3161–3168.
Cronet, P., Bellsolell, L., Sander, C., Coll, M., Serrano, L., 1995. Investigating the structural determinants of the p21-like triphosphate and Mg2+ binding site. J. Mol. Biol. 249, 654–664.
Davidi, D., Longo, L.M., Jablonska, J., Milo, R., Tawfik, D.S., 2018. A bird's-eye view of enzyme evolution: chemical, physicochemical, and physiological considerations. Chem. Rev. 118, 8786–8797.
Denessiouk, K.A., Rantanen, V.V., Johnson, M.S., 2001. Adenine recognition: a motif present in ATP-, CoA-, NAD-, NADP-, and FAD-dependent proteins. Proteins 44, 282–291.
Dishman, A.F., Tyler, R.C., Fox, J.C., Kleist, A.B., Prehoda, K.E., Babu, M.M., Peterson, F. C., Volkman, B.F., 2021. Evolution of fold switching in a metamorphic protein. Science 371, 86–90.
Dokholyan, N.V., Shakhnovich, E.I., 2001. Understanding hierarchical protein evolution from first principles. J. Mol. Biol. 312, 289–307.
Dym, O., Eisenberg, D., 2001. Sequence-structure analysis of FAD-containing proteins. Protein Sci. 10, 1712–1728.
Eck, R.V., Dayhoff, M.O., 1966. Evolution of the structure of ferredoxin based on living relics of primitive amino Acid sequences. Science 152, 363–366.
Farias-Rico, J.A., Schmidt, S., Hocker, B., 2014. Evolutionary relationship of two ancient protein superfolds. Nat. Chem. Biol. 10, 710–715.
Gifford, J.L., Walsh, M.P., Vogel, H.J., 2007. Structures and metal-ion-binding properties of the Ca2+-binding helix-loop-helix EF-hand motifs. Biochem. J. 405, 199–221.
Goncearenco, A., Berezovsky, I.N., 2010. Prototypes of elementary functional loops unravel evolutionary connections between protein functions. Bioinformatics 26, i497–i503.
Goncearenco, A., Berezovsky, I.N., 2011. Computational reconstruction of primordial prototypes of elementary functional loops in modern proteins. Bioinformatics 27, 2368–2375.
Goncearenco, A., Berezovsky, I.N., 2012. Exploring the evolution of protein function in Archaea. BMC Evol. Biol. 12, 75.
Goncearenco, A., Berezovsky, I.N., 2014. The fundamental tradeoff in genomes and proteomes of prokaryotes established by the genetic code, codon entropy, and physics of nucleic acids and proteins. Biol. Direct 9, 29.
Goncearenco, A., Berezovsky, I.N., 2015. Protein function from its emergence to diversity in contemporary proteins. Phys. Biol. 12, 045002.
Goncearenco, A., Ma, B.G., Berezovsky, I.N., 2014. Molecular mechanisms of adaptation emerging from the physics and evolution of nucleic acids and proteins. Nucleic Acids Res. 42, 2879–2892.
Gruic-Sovulj, I., Longo, L.M., Jablonska, J., Tawfik, D.S., 2022. The evolutionary history of the HUP domain. Crit. Rev. Biochem. Mol. Biol. 57, 1–15.
Guarnera, E., Berezovsky, I.N., 2016. Allosteric sites: remote control in regulation of protein activity. Curr. Opin. Struct. Biol. 37, 1–8.
Guarnera, E., Berezovsky, I.N., 2019a. On the perturbation nature of allostery: sites, mutations, and signal modulation. Curr. Opin. Struct. Biol. 56, 18–27.
Guarnera, E., Berezovsky, I.N., 2019b. Toward comprehensive allosteric control over protein activity. Structure 27, 866–878 e861.
Heizinger, L., Merkl, R., 2021. Evidence for the preferential reuse of sub-domain motifs in primordial protein folds. Proteins 89, 1167–1179.
Hocker, B., 2014. Design of proteins from smaller fragments-learning from evolution. Curr. Opin. Struct. Biol. 27, 56–62.
Holliday, G.L., Bartlett, G.J., Almonacid, D.E., O'Boyle, N.M., Murray-Rust, P., Thornton, J.M., Mitchell, J.B., 2005. MACiE: a database of enzyme reaction mechanisms. Bioinformatics 21, 4315–4316.
Holliday, G.L., Andreini, C., Fischer, J.D., Rahman, S.A., Almonacid, D.E., Williams, S.T., Pearson, W.R., 2012. MACiE: exploring the diversity of biochemical reactions. Nucleic Acids Res. 40, D783–D789.
Hua, Y.H., Wu, C.Y., Sargsyan, K., Lim, C., 2014. Sequence-motif detection of NAD(P)-binding proteins: discovery of a unique antibacterial drug target. Sci. Rep. 4, 6471.
Huang, P.S., Feldmeier, K., Parmeggiani, F., Velasco, D.A.F., Hocker, B., Baker, D., 2016. De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. Nat. Chem. Biol. 12, 29–34.
Hubbard, P.A., Liang, X., Schulz, H., Kim, J.J., 2003. The crystal structure and reaction mechanism of Escherichia coli 2,4-dienoyl-CoA reductase. J. Biol. Chem. 278, 37553–37560.
Ittah, V., Haas, E., 1995. Nonlocal interactions stabilize long range loops in the initial folding intermediates of reduced bovine pancreatic trypsin inhibitor. Biochemistry 34, 4493–4506.
Jacobs, W.M., Shakhnovich, E.I., 2017. Evidence of evolutionary selection for cotranslational folding. Proc. Natl. Acad. Sci. U. S. A. 114, 11434–11439.
Johnson, L.N., O Reilly, M., 1996. Control by phosphorylation. Curr. Opin. Struct. Biol. 6, 762–769.
Kambe, T., Taylor, K.M., Fu, D., 2021. Zinc transporters and their functional integration in mammalian cells. J. Biol. Chem. 296, 100320.
Kar, M., Posey, A.E., Dar, F., Hyman, A.A., Pappu, R.V., 2021. Glycine-rich peptides from FUS have an intrinsic ability to self-assemble into fibers and networked fibrils. Biochemistry 60, 3213–3222.
Khersonsky, O., Fleishman, S.J., 2016. Why reinvent the wheel? Building new proteins based on ready-made parts. Protein Sci. 25, 1179–1187.
King, I.C., Gleixner, J., Doyle, L., Kuzin, A., Hunt, J.F., Xiao, R., Montelione, G.T., Stoddard, B.L., DiMaio, F., Baker, D., 2015. Precise assembly of complex beta sheet topologies from de novo designed building blocks. Elife 4.
Kinoshita, K., Sadanami, K., Kidera, A., Go, N., 1999. Structural motif of phosphate-binding site common to various protein superfamilies: all-against-all structural comparison of protein-mononucleotide complexes. Protein Eng. 12, 11–14.
Koc, I., Caetano-Anolles, G., 2017. The natural history of molecular functions inferred from an extensive phylogenomic analysis of gene ontology data. PLoS One 12, e0176129.
Koczyk, G., Berezovsky, I.N., 2008. Domain Hierarchy and closed Loops (DHcL): a server for exploring hierarchy of protein domain structure. Nucleic Acids Res. 36, W239–W245.
Kolodny, R., 2021. Searching protein space for ancient sub-domain segments. Curr. Opin. Struct. Biol. 68, 105–112.
Kolodny, R., Nepomnyachiy, S., Tawfik, D.S., Ben-Tal, N., 2021. Bridging themes: short protein segments found in different architectures. Mol. Biol. Evol. 38, 2191–2208.
Lasry, I., Seo, Y.A., Ityel, H., Shalva, N., Pode-Shakked, B., Glaser, F., Berman, B., Berezovsky, I., Goncearenco, A., Klar, A., et al., 2012. A dominant negative heterozygous G87R mutation in the zinc transporter, ZnT-2 (SLC30A2), results in transient neonatal zinc deficiency. J. Biol. Chem. 287, 29348–29361.
Laurino, P., Toth-Petroczy, A., Meana-Paneda, R., Lin, W., Truhlar, D.G., Tawfik, D.S., 2016. An ancient fingerprint indicates the common ancestry of Rossmann-fold enzymes utilizing different ribose-based cofactors. PLoS Biol. 14, e1002396.
Lechner, H., Ferruz, N., Hocker, B., 2018. Strategies for designing non-natural enzymes and binders. Curr. Opin. Chem. Biol. 47, 67–76.
Lee, J., Blaber, M., 2011. Experimental support for the evolution of symmetric protein architecture from a simple peptide motif. Proc. Natl. Acad. Sci. U. S. A. 108, 126–130.
Levitt, M., Chothia, C., 1976. Structural patterns in globular proteins. Nature 261, 552–558.
Longo, L.M., Jablonska, J., Vyas, P., Kanade, M., Kolodny, R., Ben-Tal, N., Tawfik, D.S., 2020a. On the emergence of P-Loop NTPase and Rossmann enzymes from a Beta-Alpha-Beta ancestral fragment. Elife 9.
Longo, L.M., Petrovic, D., Kamerlin, S.C.L., Tawfik, D.S., 2020b. Short and simple sequences favored the emergence of N-helix phospho-ligand binding sites in the first enzymes. Proc. Natl. Acad. Sci. U. S. A. 117, 5310–5318.
Longo, L.M., Kolodny, R., McGlynn, S.E., 2022a. Evidence for the emergence of beta-trefoils by 'peptide budding' from an IgG-like beta-sandwich. PLoS Comput. Biol. 18, e1009833.
Longo, L.M., Hirai, H., McGlynn, S.E., 2022b. An evolutionary history of the CoA-binding protein Nat/Ivy. Protein Sci. 31, e4463.
Ma, B.G., Goncearenco, A., Berezovsky, I.N., 2010. Thermophilic adaptation of protein complexes inferred from proteomic homology modeling. Structure 18, 819–828.
Marcos, E., Chidyausiku, T.M., McShan, A.C., Evangelidis, T., Nerli, S., Carter, L., Nivon, L.G., Davis, A., Oberdorfer, G., Tripsianes, K., et al., 2018. De novo design of a non-local beta-sheet protein with high stability and accuracy. Nat. Struct. Mol. Biol. 25, 1028–1034.
Miller, S.L., 1953. A production of amino acids under possible primitive earth conditions. Science 117, 528–529.
Miller, S.L., Urey, H.C., 1959. Organic compound synthesis on the primitive earth. Science 130, 245–251.
Mitternacht, S., Berezovsky, I.N., 2011. Coherent conformational degrees of freedom as a structural basis for allosteric communication. PLoS Comput. Biol. 7, e1002301.
Moller, W., Amons, R., 1985. Phosphate-binding sequences in nucleotide-binding proteins. FEBS Lett. 186, 1–7.
Moret, M.A., Zebende, G.F., 2007. Amino acid hydrophobicity and accessible surface area. Phys Rev E Stat Nonlin Soft Matter Phys 75, 011920.
Mughal, F., Caetano-Anolles, G., 2023. Evolution of intrinsic disorder in protein loops. Life 13.
Murzin, A.G., Brenner, S.E., Hubbard, T., Chothia, C., 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536–540.
Narunsky, A., Kessel, A., Solan, R., Alva, V., Kolodny, R., Ben-Tal, N., 2020. On the evolution of protein-adenine binding. Proc. Natl. Acad. Sci. U. S. A. 117, 4701–4709.
Nath, N., Mitchell, J.B., Caetano-Anolles, G., 2014. The natural history of biocatalytic mechanisms. PLoS Comput. Biol. 10, e1003642.
Nepomnyachiy, S., Ben-Tal, N., Kolodny, R., 2017. Complex evolutionary footprints revealed in an analysis of reused protein segments of diverse lengths. Proc. Natl. Acad. Sci. U. S. A. 114, 11703–11708.
Newton, M.S., Guo, X., Soderholm, A., Nasvall, J., Lundstrom, P., Andersson, D.I., Selmer, M., Patrick, W.M., 2017. Structural and functional innovations in the real-

time evolution of new (betaalpha)8 barrel enzymes. Proc. Natl. Acad. Sci. U. S. A. 114, 4727–4732.
Noor, E., Flamholz, A.I., Jayaraman, V., Ross, B.L., Cohen, Y., Patrick, W.M., Gruic-Sovulj, I., Tawfik, D.S., 2022. Uniform binding and negative catalysis at the origin of enzymes. Protein Sci. 31, e4381.
Orevi, T., Rahamim, G., Hazan, G., Amir, D., Haas, E., 2013. The loop hypothesis: contribution of early formed specific non-local intreractions to determination of protein folding pathways. Biophysical Reviews 5, 85–98.
Pluckthun, A., 2015. Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. Annu. Rev. Pharmacol. Toxicol. 55, 489–511.
Qiu, K., Ben-Tal, N., Kolodny, R., 2022. Similar protein segments shared between domains of different evolutionary lineages. Protein Sci. 31, e4407.
Raanan, H., Poudel, S., Pike, D.H., Nanda, V., Falkowski, P.G., 2020. Small protein folds at the root of an ancient metabolic network. Proc. Natl. Acad. Sci. U. S. A. 117, 7193–7199.
Rigden, D.J., Galperin, M.Y., 2004. The DxDxDG motif for calcium binding: multiple structural contexts and implications for evolution. J. Mol. Biol. 343, 971–984.
Riziotis, I.G., Thornton, J.M., 2022. Capturing the geometry, function, and evolution of enzymes with 3D templates. Protein Sci. 31, e4363.
Romero Romero, M.L., Rabin, A., 2016. Tawfik DS: functional proteins from short peptides: dayhoff's hypothesis turns 50. Angew Chem. Int. Ed. Engl. 55, 15966–15971.
Romero Romero, M.L., Yang, F., Lin, Y.R., Toth-Petroczy, A., Berezovsky, I.N., Goncearenco, A., Yang, W., Wellner, A., Kumar-Deshmukh, F., Sharon, M., et al., 2018. Simple yet functional phosphate-loop proteins. Proc. Natl. Acad. Sci. U. S. A. 115, E11943–E11950.
Romero-Romero, S., Kordes, S., Michel, F., Hocker, B., 2021. Evolution, folding, and design of TIM barrels and related proteins. Curr. Opin. Struct. Biol. 68, 94–104.
Roy, S.W., Nosaka, M., de Souza, S.J., Gilbert, W., 1999. Centripetal modules and ancient introns. Gene 238, 85–91.
Samatova, E., Komar, A.A., Rodnina, M.V., 2024. How the ribosome shapes cotranslational protein folding. Curr. Opin. Struct. Biol. 84, 102740.
Saraste, M., Sibbald, P.R., Wittinghofer, A., 1990. The P-loop–a common motif in ATP- and GTP-binding proteins. Trends Biochem. Sci. 15, 430–434.
Sato, Y., Niimura, Y., Yura, K., Go, M., 1999. Module-intron correlation and intron sliding in family F/10 xylanase genes. Gene 238, 93–101.
Schaeffer, R.D., Kinch, L.N., Liao, Y., Grishin, N.V., 2016. Classification of proteins with shared motifs and internal repeats in the ECOD database. Protein Sci. 25, 1188–1203.
Schimmel, P.R., Flory, P.J., 1967. Conformational energy and configurational statistics of poly-L-proline. Proc. Natl. Acad. Sci. U. S. A. 58, 52–59.
Schneider, G., Neuberger, G., Wildpaner, M., Tian, S., Berezovsky, I., Eisenhaber, F., 2006. Application of a sensitive collection heuristic for very large protein families: evolutionary relationship between adipose triglyceride lipase (ATGL) and classic mammalian lipases. BMC Bioinf. 7, 164.
Seal, M., Weil-Ktorza, O., Despotovic, D., Tawfik, D.S., Levy, Y., Metanis, N., Longo, L.M., Goldfarb, D., 2022. Peptide-RNA coacervates as a cradle for the evolution of folded domains. J. Am. Chem. Soc. 144, 14150–14160.
Shimada, J., Yamakawa, H., 1984. Ring-closure probabilities for twisted wormlike chains. Application to DNA. Macromolecules 17, 689–698.
Siddiq, M.A., Hochberg, G.K., Thornton, J.W., 2017. Evolution of protein specificity: insights from ancestral protein reconstruction. Curr. Opin. Struct. Biol. 47, 113–122.
Smith, C.A., Rayment, I., 1996. Active site comparisons highlight structural similarities between myosin and other P-loop proteins. Biophys. J. 70, 1590–1602.
Smock, R.G., Yadid, I., Dym, O., Clarke, J., Tawfik, D.S., 2016. De novo evolutionary emergence of a symmetrical protein is shaped by folding constraints. Cell 164, 476–486.
Sterner, R., Höcker, B., 2005. Catalytic versatility, stability, and evolution of the (betaalpha)8-barrel enzyme fold. Chem. Rev. 105, 4038–4055.
Taylor, T.J., Vaisman II, 2006. Graph theoretic properties of networks formed by the Delaunay tessellation of protein structures. Phys Rev E Stat Nonlin Soft Matter Phys 73, 041925.
Tee, W.V., Guarnera, E., Berezovsky, I.N., 2020. Disorder driven allosteric control of protein activity. Current Research in Structural Biology 2, 191–203.
Tee, W.V., Tan, Z.W., Lee, K., Guarnera, E., Berezovsky, I.N., 2021. Exploring the allosteric territory of protein function. J. Phys. Chem. B 125, 3763–3780.
Tee, W.V., Tan, Z.W., Guarnera, E., Berezovsky, I.N., 2022. Conservation and diversity in allosteric fingerprints of proteins for evolutionary-inspired engineering and design. J. Mol. Biol. 167577.
Thommen, M., Holtkamp, W., Rodnina, M.V., 2017. Co-translational protein folding: progress and methods. Curr. Opin. Struct. Biol. 42, 83–89.
Traut, T.W., 1994. The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. Eur. J. Biochem. 222, 9–19.
Trifonov, E.N., 1999. Glycine clock: eubacteria first, Archaea next, protoctista, fungi, planta and animalia at last. Gene Ther. Mol. Biol. 4, 313–322.
Trifonov, E.N., 2000. Consensus temporal order of amino acids and evolution of the triplet code. Gene 261, 139–151.
Trifonov, E.N., Berezovsky, I.N., 2002. Molecular evolution from abiotic scratch. FEBS (Fed. Eur. Biochem. Soc.) Lett. 527, 1–4.
Trifonov, E.N., Berezovsky, I.N., 2003. Evolutionary aspects of protein structure and folding. Curr. Opin. Struct. Biol. 13, 110–114.
Trifonov, E.N., Kirzhner, A., Kirzhner, V.M., Berezovsky, I.N., 2001. Distinct stages of protein evolution as suggested by protein sequence analysis. J. Mol. Evol. 53, 394–401.
Trudeau, D.L., Kaltenbach, M., Tawfik, D.S., 2016. On the potential origins of the high stability of reconstructed ancestral proteins. Mol. Biol. Evol. 33, 2633–2641.
van der Gulik, P., Massar, S., Gilis, D., Buhrman, H., Rooman, M., 2009. The first peptides: the evolutionary transition between prebiotic amino acids and early proteins. J. Theor. Biol. 261, 531–539.
Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al., 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 50, D439–D444.
Vyas, P., Trofimyuk, O., Longo, L.M., Deshmukh, F.K., Sharon, M., Tawfik, D.S., 2021. Helicase-like functions in phosphate loop containing beta-alpha polypeptides. Proc. Natl. Acad. Sci. U. S. A. 118.
Vyas, P., Malitsky, S., Itkin, M., Tawfik, D.S., 2023. On the origins of enzymes: phosphate-binding polypeptides mediate phosphoryl transfer to synthesize adenosine triphosphate. J. Am. Chem. Soc. 145, 8344–8354.
Walker, J.E., Saraste, M., Runswick, M.J., Gay, N.J., 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1, 945–951.
Xie, X., Backman, D., Lebedev, A.T., Artaev, V.B., Jiang, L., Ilag, L.L., Zubarev, R.A., 2015. Primordial soup was edible: abiotically produced Miller-Urey mixture supports bacterial growth. Sci. Rep. 5, 14338.
Yamakawa, H., Stokmayer, W.H., 1972. Statistical mechanics ofwormlike chains. 2. Excluded volume effects. J Chem Phys Biol 57, 2843–2854.
Yew, B.K., Chintapalli, S.V., Upton, G.G., Reynolds, C.A., 2007. Conservation of closed loops. J. Mol. Graph. Model. 26, 652–655.
Yin, M., Goncearenco, A., Berezovsky, I.N., 2021. Deriving and using descriptors of elementary functions in rational protein design. Front Bioinform 1, 657529.
Zeldovich, K.B., Berezovsky, I.N., Shakhnovich, E.I., 2006. Physical origins of protein superfamilies. J. Mol. Biol. 357, 1335–1343.
Zheng, Z., Goncearenco, A., Berezovsky, I.N., 2016. Nucleotide binding database NBDB–a collection of sequence motifs with specific protein-ligand interactions. Nucleic Acids Res. 44, D301–D307.